PRINCE OF SONGKLA UNIVERSITY
FACULTY OF ENGINEERING

Final Examination: Semester 2 **Academic Year**: 2014

Date: 11 May 2015 **Time**: 9.00 – 12.00 (3 hours)

Subject Number: 242-441 **Room**: S103

Subject Title: Advanced Computer Architecture and Organization

---

Exam Duration: 3 hours

This paper has 14 pages, 3 questions and 180 marks (25%).

Authorised Materials:

- Writing instruments (e.g. pens, pencils).
- Textbooks, a notebook, handouts, and dictionaries are permitted.

Instructions to Students:

- Scan all the questions before answering so that you can manage your time better.
- Answers **must** be written in **Thai**.
- Write your name and ID on every page.
- Any unreadable parts will be considered wrong.

When drawing diagrams or coding, use good layout, and short comments; marks will not be deducted for minor syntax errors.

Cheating in this examination

Lowest punishment: Failed in this subject and courses dropped for next two semesters.

Highest punishment: Expelled.

NO	Time (Min)	Marks	Collected
1	65	65	
2	95	95	
3	20	20	
Total	180	180	
100%		25%	

Question 1 Multiprocessors and Thread-Level Parallelism (65 marks; 65 minutes)

1.1 Tell whether the following statements are true (T) or false (F). (25 marks)

a) _____ Centralized Memory Multiprocessors also are called symmetric multiprocessors (SMPs) because single main memory has a symmetric relationship to all processors.

b) _____ Centralized Memory Multiprocessors have large caches, so that single memory can satisfy memory demands of a small number of processors.

c) _____ In Distributed Memory Multiprocessors, communicating data between processors more complex.

d) _____ In Non-Uniformed Memory Access (NUMA), all processes can access all memory modules using the same amount of time.

e) _____ In Uniformed Memory Access (UMA), each processor has a local memory.

f) _____ In Distributed Shared-Memory (DSM), Cache controller can simply snoop on a shared memory bus.

g) _____ Examples of architectures with Non-Uniformed Memory Access (NUMA) are Parallel Vector Processor (PVP) and Symmetric Multiprocessors (SMP).

h) _____ In Vector Processors, the instruction set includes operations on vectors and also scalars.

i) _____ In Processing Array, the CPU speed increases when conditionally executing code.

j) _____ Processor Arrays naturally supports multiple users.

k) _____ In Symmetric Multiprocessors, the same address on different CPUs refers to different memory locations.

l) _____ In Symmetric Multiprocessors, processors communicate via shared data values.

m) _____ In Symmetric Multiprocessors, memory access time same for all CPUs.

n) _____ Replicating reduces contention among processors for shared data values but CPUs may have obsolete images of address locations stored in their cache.

o) _____ Example interconnections of Massively Parallel Processors (MPPs) are hypercube and mesh.

p) _____ Massively Parallel Processors (MPPs) communicate data using shared memory.

q) _____ In Cache Coherence, to write a value, the processor must have an exclusive access to that address location first.

r) _____ In Cache Coherence, before writing, data values in other caches will be validated.

s) _____ In Cluster of Workstations (COW), each node is a virtual machine.

t) _____ In Asymmetrical Cluster, the front end can become a single point for

failure.

u) _____ In Asymmetrical Cluster, the performance capabilities of the front end computer limits the system's scalability.

v) _____ In Asymmetrical Cluster, every computer executes the same OS and has identical functionality and users can login to any computer to edit or compile the programs.

w) _____ In Symmetrical Cluster, CPU cycles are dedicated to parallel computing.

x) _____ We can reduce the frequency of remote accesses that affects parallel processing either by caching shared data in hardware or restructuring the data layout in software to make more accesses local.

y) _____ In Snooping Protocol, complexity from retrieving cache block from a processor cache, which can take longer than retrieving it from memory.

1.2 Compare the following items. (6 marks)

a) Centralized Memory Multiprocessor and Physically Distributed-Memory multiprocessor (4 marks)

b) Pipelined Vector Processor and Processor Array (2 marks)

1.3 Fill in the space. (12 marks)

a) In _____, each CPU cache's controller will monitor the bus to see which cache blocks are being requested by other CPUs.

b) _____ is a situation in which at most one process can be engaged in a specified activity at any time.

c) _____ guarantees that no process will proceed beyond a designated point in the program, until every process has reached that point.

d) _____ is a large-scale distributed memory system with many individual nodes

e) In _____, the system hardware and software create an illusion of a single address space to users.

f) In Distributed Shared-Memory (DSM), _____ is used to support distributed coherent caches.

g) In Distributed Shared-Memory (DSM), _____ directory entry/ies is assigned for each cache block.

h) In Asymmetrical Cluster, _____ interacts with users and IO devices while _____ are dedicated to executing parallel programs.

i) In Symmetric Shared-Memory Multiprocessors, there is the 4th cache miss apart from Compulsory, Capacity, and Conflict misses which is _____ _miss due to _____ when attempts to writing data that has been copied to other cache blocks.

j) In _____ Cache Coherence Protocol, every memory block has associated directory information that keeps track of copies of cached blocks and their states.

1.4 Calculate the following answers. (6 marks)

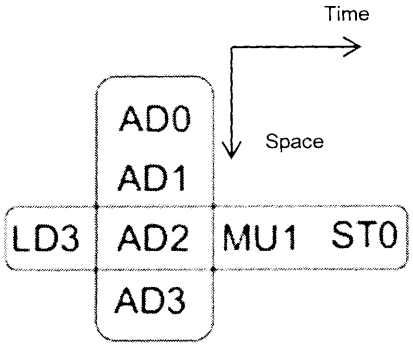
a) Suppose that we expect 100x Speedup with 256 cores. How many percentage of the original program can be sequential? (3 marks)

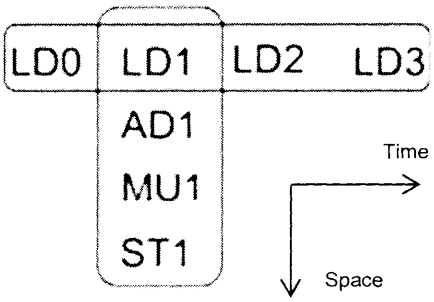
b) Suppose 32 CPU MP, 2GHz, 200 ns remote memory, all local accesses hit memory hierarchy and base CPI is 0.5. (Remote access = 200/0.5 = 400 clock cycles.) What is performance impact if 0.2% instructions involve remote access? (3 marks)

1.5 Explain the followings. (16 Marks)

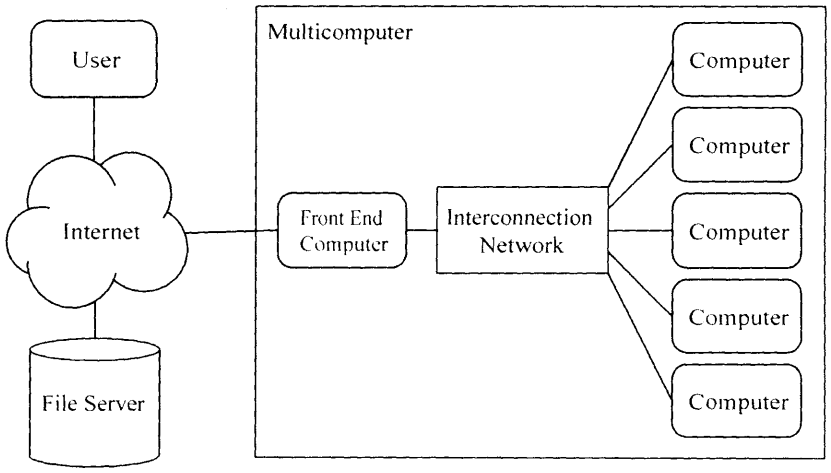
a) Explain problems associated with shared data. (4 marks)

b) Tell which of the following pictures show how instructions are processed by Pipelined Vector Processor or Processor Array. (4 marks)

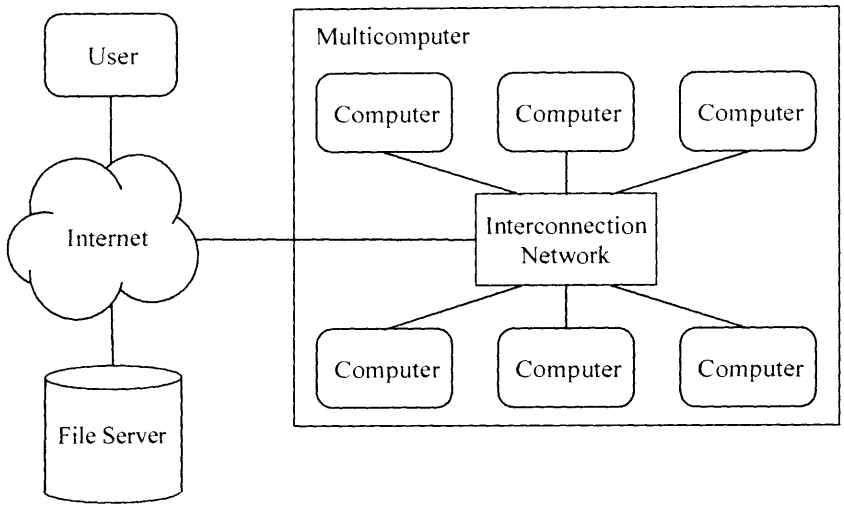




c) From the following pictures, tell which are an Asymmetrical Cluster or a Symmetrical Cluster (2 marks)



Name_____ ID_____



d) Explain the following cache block sharing status in Directory Based Cache Coherence Protocol. (6 marks)

Uncached:

Shared:

Exclusive:

Question 2 Parallel Computing, Performance Analysis and Load Balancing (95 marks; 95 minutes)

2.1 Answer the following questions. (40 marks, 40 minutes)

a) List 4 **hardware factors** that play a significant role in scalability. (4 marks)

b) What are factors that contribute to **parallel overhead**? (4 marks)

c) Compare *loop independent data dependence* and *loop carried data dependence* by giving an example of code fragment for each type. (4 marks)

Loop independent data dependence	Loop carried data dependence

d) Give an example of problems that result in load imbalances even if data is evenly distributed among tasks. (2 marks)

e) List at least 4 factors to consider when designing your program's inter-task communications. (4 marks)

f) What are *Dynamic Load Balancing Factors?* (6 marks)

g) Explain at least 3 approaches concerning *System Information exchange policy?* (6 marks)

h) Explain 7 *Properties* of Load Balancing Systems. (7 marks)

i) How to increase throughput in parallel computing? (2 marks)

j) What are the importance parameters for *Memory Performance* and how to measure them? (3 marks)

2.2 Tell whether the following statements are true (T) or false (F). (25 marks)

a) _____ Coarse-grain Parallelism has a high computation to communication ratio that facilitates load balancing.

b) _____ Fine-grain Parallelism has a low computation to communication ratio that facilitates load balancing.

c) _____ Fine-grain Parallelism implies low communication overhead and more opportunity for performance enhancement.

d) _____ In Fine-grain Parallelism, it is harder to load balance efficiently.

e) _____ Load balancing refers to the practice of distributing work among tasks so that all tasks are kept busy all of the time.

f) _____ If all tasks are subject to a barrier synchronization point, the slowest task will determine the overall performance.

g) _____ The strategy is to focus on parallelizing the hotspots and ignore those sections of the program that account for little CPU usage.

h) _____ Bottlenecks in the program are in the areas that are disproportionately slow, or cause parallelizable work to halt or be deferred.

i) _____ When bottlenecks are found, restructuring the program or using a different algorithm will help reduce or eliminate unnecessary slow areas.

j) _____ A common inhibitor in parallel computing is data dependence.

k) _____ Load balancing strategies try to migrate tasks from less loaded machines to heavily loaded ones.

l) _____ The load migration has to maximize the response time and optimize the overall system performance.

m) _____ Load balancing concerns scheduling or resource allocation and management.

n) _____ Migration of a blocked process is useful because it affects the local processor load.

o) _____ Smaller processes put more load on the communication network.

p) _____ Migrating the process with the highest remaining service time will benefit most in the long-term.

q) _____ Processes which communicate frequently with the intended destination processors, will reduce communications load if they migrate.

r) _____ Migrating the most locally demanding process will be of great benefit to local load reduction.

s) _____ A good factor for measuring the cost-effectiveness is utilization.

t) _____ Isoefficiency is a way to measure scalability.

u) _____ A scalable system maintains efficiency as processors are added or the problem size increases.

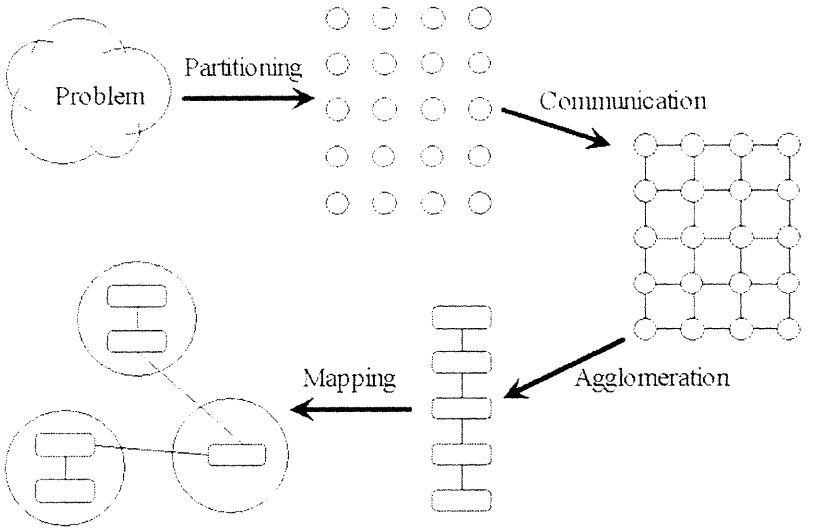
v) _____ A system with small Isoefficiency function is a system that cannot scale well.

w) _____ We can keep the speedup fixed by increasing both the size of problem and number of processors.

x) _____ System throughput is the ratio of the achieved speed to the peak speed of a given computer.

y) _____ Utilization is a number of jobs processed in a unit time.

2.3 Use the following figure to explain what need to be done in the design methodology in Question c) – j). (7 marks)



a) Partitioning (1 mark)

b) Communication (1 mark)

c) Agglomeration (1 mark)

d) Granularity (1 mark)

e) Mapping (1 mark)

f) What are the conflicting goals of mapping? (2 marks)

2.4 From the following list, put the matched items to the design methodology check lists. (13 marks)

A. Minimize redundant computations and redundant data storage

B. Communication operations balanced among tasks

C. Locality of parallel algorithm has increased

D. Each task communicates with only small group of neighbors

E. Replicated computations take less time than communications they replace

F. Number of tasks an increasing function of problem size

G. Data replication doesn't affect scalability

H. Tasks can perform communications concurrently

I. Primitive tasks roughly the same size

J. Number of tasks increases with problem size

K. Considered designs based on one task per processor and multiple tasks per processor

L. Number of tasks suitable for likely target systems

M. Evaluated static and dynamic task allocation

a) Partitioning

---

b) Communication

---

c) Agglomeration

---

d) Mapping

2.5 Draw graphs that illustrate the following items of Amdahl Laws. Give details of all axes and legends. (10 marks)

a) Typical plot for showing the scalability of a parallel system (speedup changes as the problem size and the number of processors change). (4 marks)

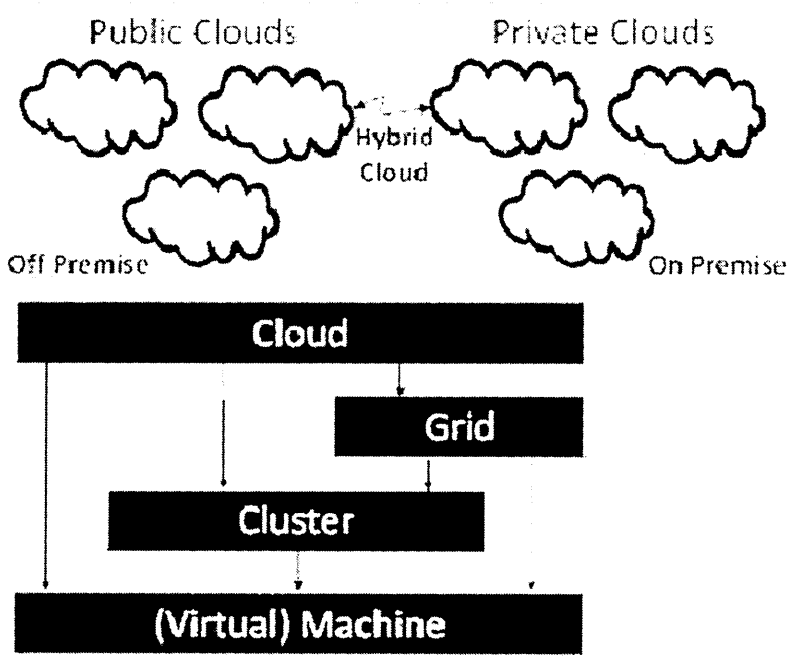
b) Typical efficiency plot for a fixed problem size. (3 marks)

c) Typical efficiency plot for a fixed problem size. (3 marks)

Question 3 Grid Technology and Cloud Computing (20 marks; 20 minutes)

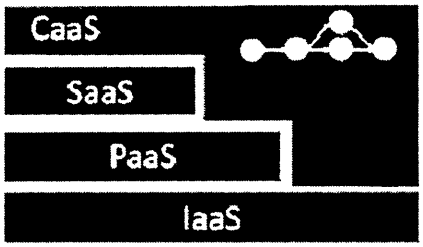
Answer the following questions.

a) Describe and compare all keywords in the following figure and their relationship. (10 marks)



Name_____ ID_____

b) Explain and compare all keywords in the following figure when Composite as a Service (CaaS) includes Composite Services and composes a workflow that links and orchestrates distributed applications over the Internet. Also give some examples of such services. (10 marks)



----End of Examination----

Pichaya Tandayya Lecturer

Name_____ ID_____